BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 9: e181431
doi: 10.3897/biss.9.181431

OPEN ACCESS

Conference Abstract

Modelling Complex Marine Survey Data Using the Darwin Core Data Package: The BROKE-West Fish Dataset

Yi-Ming Gan[‡], John Richard Wieczorek[§], Anton Pieter Van de Putte[‡,|], Cecilie Svenningsen[¶], Pablo Deschepper[‡], Charlie Plasman[‡], Tim Robertson[¶]

‡ Royal Belgian Institute of Natural Sciences, Brussels, Belgium
§ Rauthiflor LLC, Bariloche, Argentina
| Université Libre de Bruxelles, Brussels, Belgium
¶ Global Biodiversity Information Facility, Copenhagen, Denmark

Corresponding author: Yi-Ming Gan (ymgan@naturalsciences.be)

Received: 05 Dec 2025 | Published: 08 Dec 2025

Citation: Gan Y-M, Wieczorek J, Van de Putte A, Svenningsen C, Deschepper P, Plasman C, Robertson T (2025) Modelling Complex Marine Survey Data Using the Darwin Core Data Package: The BROKE-West Fish Dataset. Biodiversity Information Science and Standards 9: e181431. https://doi.org/10.3897/biss.9.181431

Abstract

The BROKE-West fish dataset (Australian Antarctic Data Centre 2022) originates from the BROKE-West voyage, a survey conducted in a data-poor area of the Southern Ocean in 2006. This dataset forms part of the baseline biodiversity assessment of the Cosmonaut and Cooperation Seas.

The dataset was initially published in 2008 as a single flat Darwin Core Archive (DwC-A) occurrence table without extensions. Although the original dataset contained valuable occurrence data, it omitted much of the sampling context, including survey design, target scope, and material relationships. In 2016, the dataset was reused to test the extended MeasurementOrFact (eMoF) extension developed by Ocean Biodiversity Information System (OBIS) (De Pooter et al. 2017), which enabled the inclusion of environmental and biological measurements (Van de Putte 2016). In 2023, the Humboldt Ecological Inventory extension (TDWG Humboldt Extension Task Group and Biodiversity Information Standards (TDWG) 2024) was applied to describe the survey design and scope, linking each trawl to its survey targets and sampling context (Van de Putte et al. 2025).

© Gan Y et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The presentation from which this extended abstract derives (Gan et al. 2025) describes the limitations of DwC-A for this dataset and how DwC-DP overcomes these.

While DwC-A remains useful, several concepts cannot be well represented in this format.

1. Marine datasets often contain both community-level measurements (e.g., abundance or total weight of an occurrence) and individual-level measurements (e.g., the standard length of a single fish from that same occurrence). The subsetting relationship is not explicit in DwC-A.
2. Survey context and survey target are flattened into a single Humboldt Extension table. If each survey target combination is recorded as a separate row, survey context information, such as *eco:isTaxonomicScopeFullyReported* of a *dwc:Event*, must be repeated for each target.
3. The use of identifiers can be confusing: because the Occurrence extension requires a unique *dwc:occurrenceID* per row, the sample identifier representing a dwc:MaterialEntity must be used as dwc:occurrenceID. This leads to properties of the material (e.g., dry weight) being linked via eMoF to an Occurrence instead of to the material itself.
4. Although photographs can be linked using *dwc:associatedMedia*, this approach omits key media metadata, such as the license.
5. Prey items found in the stomach contents of fish from an occurrence are difficult to model in DwC-A.

Today, the dataset has been fully restructured (Gan and Van de Putte 2025) using the proposed Darwin Core Data Package (DwC-DP) (Darwin Core Maintenance Group 2025). DwC-DP replaces the single-core architecture with a linked-table model that overcomes the limitations of the star schema. This new structure allows an explicit representation of *Occurrence* and *Material*, preserving the meaning of *dwc:occurrenceID* and *dwc:materialEntityID*. It also supports explicit relationships such as *isPartOf* and *derivedFrom*. For example, a prey item with *dwc:materialEntityID* “AAV3FF_00337_stomach_002_Euphausiid” is *dwc:derivedFromMaterialEntityID* “AAV3FF_00337” with *dwc:derivationType* “stomach content of”.

This restructuring has revealed the full analytical potential of legacy datasets. The BROKE-West case demonstrates how DwC-DP can represent data that were previously unpublished or too complex to format in DwC-A, such as material-to-material derivations and media linked to materials.

The following three examples highlight DwC-DP’s analytical flexibility.

1. Non-detections for an event can be inferred by comparing occurrences to survey targets. If all occurrences of a survey target are fully reported, an event without that survey target indicates a non-detection.
2. Fish length-frequency distributions can be computed directly by linking *Material Assertion* to *Material* and then to *Event*, aligning with the Essential Ocean

Variable (EOV) sub-variable Fish Abundance and Distribution (Global Ocean Observing System 2025).

3. Diet composition can be represented through material relationships between predator stomachs and prey taxa — an analysis previously impossible in DwC-A because stomach-content data were unpublished. An OrganismInteraction table was not used since many prey items were unidentifiable or digested and their ingestion time unknown, making the creation of corresponding prey Occurrences difficult. These examples underline how DwC-DP enhances both data representation and the interpretability and reusability of historical datasets.

Adopting DwC-DP brings major benefits but also challenges. It better preserves data meaning, supports richer links between entities, and suits complex marine survey datasets. However, it requires more identifiers, more relational workflows, and careful curation to keep identifiers stable over time.

DwC-DP is a major shift toward a relational biodiversity data model that accommodates complex ecological surveys. The BROKE-West case study shows how DwC-DP enriches flat datasets and more accurately represents survey complexity.

Keywords

DwC-DP, biodiversity data, data standard, data sharing

Presenting author

Yi-Ming Gan

Presented at

Living Data 2025

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Australian Antarctic Data Centre (2022) RMT Trawl catch from the 2005/06 V3 BROKE-West voyage - Fish. https://doi.org/10.15468/nooqmj. Accessed on: 2025-5-12.
- Darwin Core Maintenance Group (2025) Darwin Core Data Package guide. Biodiversity Information Standards (TDWG). http://rs.tdwg.org/dwc/doc/dp/2025-09-10

- De Pooter D, Appeltans W, Bailly N, et al. (2017) Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. Biodiversity Data Journal 5 https://doi.org/10.3897/bdj.5.e10989
- Gan Y, Van de Putte A (2025) Distribution of squid and fish in the pelagic zone of the Cosmonaut Sea and Prydz Bay region during the BROKE-West campaign - data. 18. Release date: 2025-11-25. URL: https://doi.org/10.5281/zenodo.17708634
- Gan Y, Wieczorek J, Svenningsen C, et al. (2025) Modelling Complex Marine Survey Data Using the Darwin Core Data Package. Presentation. https://doi.org/10.5281/zenodo.17701152
- Global Ocean Observing System (2025) Essential Ocean Variable Specification Sheet: Fish abundance and distribution. https://goosocean.org/document/17510. Accessed on: 2025-10-14.
- TDWG Humboldt Extension Task Group, Biodiversity Information Standards (TDWG) (2024) Humboldt Extension Vocabulary List of Terms. http://rs.tdwg.org/dwc/doc/eco/2024-03-26. Accessed on: 2025-9-09.
- Van de Putte A (2016) AntOBIS: Fish length and distribution during the BROKE-West survey. . 1.11. No Organisation. Release date: 2016-11-03. URL: https://ipt.iobis.org/obis-env/resource?r=brokewest_fish&v=1.3
- Van de Putte A, Jackson G, Pakhomov E, et al. (2025) Distribution of squid and fish in the pelagic zone of the Cosmonaut Sea and Prydz Bay region during the BROKE-West campaign - data. 1.11. Test Organization #1. Release date: 2025-10-09. URL: https://ipt.gbif.org/resource?r=brokewest-fish&v=1.11